AD-A274 148

8 copy p. 42

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AEOSR-TR-

6a. NAME OF PERFORMING ORGANIZATION
Department of Psychology
SUNY Stony Brook

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code)
Stony Brook, NY 11794-2500

7b. ADDRESS (City, State and ZIP Code)
Bolling Air Force Base, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AFOSR

8b. OFFICE SYMBOL (If applicable)
NL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR 91-0378

8c. ADDRESS (City, State and ZIP Code)
110 Duncan ave., Suite B115
Bolling AFB DC 20332-0001

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	AS	

11. TITLE (Include Security Classification)
Signal- and Listener-based Factors in Complex Auditory ~~Pattern Perception~~

12. PERSONAL AUTHOR(S)
Arthur G. Samuel

13a. TYPE OF REPORT
Annual Technical

13b. TIME COVERED
FROM 9/15/92 TO 9/15/93

14. DATE OF REPORT (Yr., Mo., Day)
1993-11-9

15. PAGE COUNT
13

16. SUPPLEMENTARY NOTATION

DTIC ELECTE DEC 27 1993 S A

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The research conducted during the second year of AFOSR grant # 91-0378 investigated fundamental issues in the early processing of speech and similarly complex acoustic signals. The research pursued the information processing goal of specifying the levels of analysis that occur between the initial sensory coding of the signal, and the recognition of the phonetic sequence it conveys. Five experiments provided evidence for the existence of at least three qualitatively different levels of perceptual analysis. The data help to specify the properties of each level, including a locus (peripheral vs. central), a stimulus domain, and the mechanisms affected by repeated stimulation. The convergence across several different approaches used to determine levels of analysis supports the three-level model.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. John Tangney

22b. TELEPHONE NUMBER (Include Area Code)
202-767-5021

22c. OFFICE SYMBOL
AFOSR/NL

DD FORM 1473, 83 APR

EDITION OF 1

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

93 12 23 082

93-31237

I. SUMMARY

The research conducted during the second year of AFOSR grant # 91-0378 investigated fundamental issues in the early processing of speech and similarly complex acoustic signals. The research pursued the information processing goal of specifying the levels of analysis that occur between the initial sensory coding of the signal, and the recognition of the phonetic sequence it conveys. Five experiments provided evidence for the existence of at least three qualitatively different levels of perceptual analysis. The data help to specify the properties of each level, including a locus (peripheral vs. central), a stimulus domain, and the mechanisms affected by repeated stimulation. The convergence across several different approaches used to determine levels of analysis supports the three-level model.

DTIC QUALITY INSPECTED 5

Accesion For
NTIS CRA&I
DTIC TAB
Unannounced
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1

II. RESEARCH OBJECTIVES

The objective of the research project is to delineate principles that underlie the perception of complex auditory patterns. The stimuli used are speech and musical patterns of varying complexity. A wide array of experimental procedures and analyses are used to try to determine properties that are true of the perception of complex auditory patterns across stimulus domains. In addition, we also are interested in discovering any principles that are domain specific (e.g., as "categorical perception" has traditionally been claimed to be a principle of perception specific to the speech domain). The various experimental investigations in the project may be broadly grouped into studies of signal-based factors, and studies of listener-based factors. The former group includes experiments that explore how properties of the input signal determine perception, while the latter group includes studies of how listeners' expectations influence perception/performance. The former group primarily focusses on early representations of the signal, and the latter includes higher-level factors (including, but not limited to, attentional influences). The long-term goal of the research is to understand both signal-based and listener-based factors, and their interaction in the perception of complex auditory patterns.

III. PROGRESS TOWARD RESEARCH OBJECTIVES

The second year of funding for AFOSR 91-0378 has been very productive. During this time, we have brought our new laboratory at Stony Brook on-line, after an initial year of transition. During this past year, we have been able to collect data, analyze data, and produce manuscripts using our new facilities, all at a more productive level than was possible before the establishment of our new laboratory.

During this period, we wrote and submitted three major manuscripts to leading journals. Two of these were based on research that was summarized in last year's Annual Technical Report. The current report will summarize the research contained in the third current manuscript. Additional research being conducted in collaboration with Lee Wurm, a graduate trainee, will be summarized separately in our Annual Technical Report for AASERT grant # 93NL174. Collectively, the various lines of research pursued this year have provided significant advances in our understanding of the representations and processes involved in the perception of speech and other complex sounds.

A fundamental premise of the information processing perspective is that perceptual and cognitive functioning may usefully be decomposed into levels of analysis. For understanding how complex sounds are perceived, this perspective entails providing a specification of what each level of analysis is, and the relationship of each level to other levels. In specifying a level of analysis, there are many kinds of information that we should want to know. For example, it is important to delineate the domain of operation: Does a process at a particular level of analysis only operate, say, on auditory stimuli, or perhaps only on auditory signals from one ear, or only on signals with certain properties (e.g., musical sounds), etc. To the extent that we can specify the stimulus properties that are critical to an analysis at a given level, we understand the nature of the system. Moreover, we must understand the mechanisms of processors at a given level. Do they, for example, change their output as a function of the stimulation, or are they stable over time? Does the activation of a particular representation have any effect on other representations at the same level, or is each one independent? Similarly, we should try to understand how the activity at one level of the system affects the behavior of processors at other levels.

The research conducted this year has allowed us to identify a number of qualitatively separable levels of analysis that complex sounds undergo. For each hypothesized level of analysis, we have evidence for its domain of operation, and for a number of its operating characteristics. For example, the data indicate whether a given process is monaurally-driven or binaurally-driven, and whether it is sensitive to properties that are directly specified in a "neural spectrogram", or is instead sensitive to more abstract, derived properties. Separate levels of analysis can be inferred when processors at two putative levels show consistently different sets of operating principles.

There are at least three literatures that have used this general approach in the study of speech and other complex sounds. Fujisaki and Kawashima (1969, 1970; Pisoni, 1973) used qualitatively different patterns of discrimination performance over time to argue for the existence of two levels of analysis. One of these was characterized as "acoustic", and the other "phonetic". Representations at the acoustic level were posited to preserve most of the stimulus detail, but were subject to decay over the course of a second or so, whereas phonetic codes were more abstract (i.e., contained less of the original information) but were more stable over time. The Fujisaki and Kawashima model used the concept of two qualitatively different levels to account for patterns of categorical perception, and helped to tie together the results for variations in stimulus type (vowels versus consonants) and testing conditions (timing manipulations in the discrimination tests).

Using a quite different phenomenon, Cutting (1976) argued for multiple levels of analysis for speech. Rather than categorical perception, Cutting focussed on various types of dichotic fusion. For example, under some conditions, presenting /ba/ to one ear simultaneously with /ga/ in the other will produce a percept of /da/; Cutting called this type of dichotic fusion "psychoacoustic fusion", as it appears to be due to some sort of averaging of /ba/'s rising second formant with /ga/'s sharply falling one to produce the relatively flat F2 of /da/. Cutting contrasted this type of fusion with a case such as /ba/ paired with /ta/ yielding a percept of /da/ or /pa/. In this case, it appears that the voicing and place features of the input get recombined, hence the label of "phonetic feature fusion". Cutting looked at six different types of fusion, and showed that at least three different levels of analysis were needed to account for how the fusions behaved as a function of testing conditions (variations in dichotic onset times, intensity, and frequency). As in the categorical perception literature, patterns of performance were most parsimoniously accounted for by positing the existence of multiple levels of analysis.

Our research continues a third approach that has been used to establish processing levels, based on a third phenomenon: selective adaptation. In the adaptation paradigm, listeners identify stimuli forming a continuum, for example, from /da/ to /ta/. Eimas and Corbit (1973) first demonstrated that repeated presentation of one of the continuum endpoints causes listeners to report fewer stimuli as members of the adaptor's category. In addition, they showed that the adaptation effect could be obtained with adaptors that were not members of the test continuum, if the adaptors shared important properties with the test items. For example, Eimas and Corbit reported that repeatedly presenting /ba/ reduced report of /da/ in a /da/-/ta/ continuum in much the same way that /da/ itself did. They argued that this result indicated that the adaptation effect could occur at a phonetic feature level of analysis, the level shared by the voiced /b/ and /d/.

In the years since Eimas and Corbit's (1973) seminal paper, several investigators have argued that multiple levels of analysis are needed to account for the pattern of adaptation effects. Tartter and Eimas (1975) first suggested that adaptation shifts could occur at the levels posited by Fujisaki and Kawashima (1969, 1970): In addition to the phonetic level implicated by the original Eimas and Corbit study, Tartter and Eimas argued that a more acoustic level of representation was needed to account for their observation of adaptation effects when the adaptors were (non-phonetic) single formants. Using evidence from several different approaches, a number of other researchers argued that the data require two levels (e.g., Kat and Samuel, 1984; Samuel, 1986; Samuel and Newport, 1979; Sawusch, 1977), or possibly three (Sawusch, 1986). In the past year, we conducted an extensive set of adaptation experiments to investigate the issue of levels of representation. These studies have proven very fruitful, and have allowed us to develop a three-level information processing model. Table 1 summarizes this model.

Insert Table 1 Here

The model posits three different levels of representation. A level is justified tc the extent that we can show that it has qualitatively different properties than those of another posited level. For example, one way that Level I is differentiated from Levels II and III is in terms of its locus; this level of representation consists of monaurally-driven units, whereas Levels II and III are binaurally driven. Evidence for such monaurally-driven representations can be found in a number of studies, including those by Sawusch (1977) and by Samuel (1988). In Sawusch's study, subjects identified stimuli from a /bae-dae/ continuum. The adaptors included the endpoints of the continuum, and stimuli made by modifying the formant patterns of these endpoints. These modified adaptors were synthesized using formant tracks with the same shape as the endpoints, but each formant was shifted up the frequency scale by 1.5 critical bands. Sawusch found that these shifted adaptors produced a reliable adaptation effect on /bae-dae/ identification, of about half the size of the effect of the endpoints. Most interestingly, when the adaptors and test syllables were presented monaurally, the shifted adaptors were just as effective when they were presented contralaterally to the test syllables, while the endpoint adaptors lost half of their efficacy when they were presented contralaterally, rather than ipsilaterally. The larger ipsilateral effect for the endpoint adaptors implicates a monaurally-driven component to the effect. The same conclusion was supported by Samuel's (1988) study, using a /ba-wa/ test series. Samuel found that the endpoint /wa/ was approximately twice as effective under ipsilateral adaptation conditions as under contralateral conditions, again supporting a monaural component. These studies (and others) also implicate a binaurally-driven component, because of the reliable (though reduced) adaptation effects found for the contralateral tests.

Our recent experiments have developed evidence to support a distinction between two additional levels of representation. These experiments have used the laterality manipulation used by Sawusch (1977) and Samuel (1988), together with two additional methodological tools. One of these tools, like Sawusch's formant-shifted adaptors, is stimulus-based, while the second is procedural/analytic. The test stimuli in these experiments are members of a /ba-da/ continuum. We have conducted a number of experiments using three types of adapting stimuli: (1) endpoint adaptors, (2) adaptors sharing the endpoints' second and/or third formant patterns, but missing F1, and (3) adaptors sharing phonetic properties with the endpoints, but mismatching acoustically (/pa/ and /ta/). In all of these experiments subjects are instructed to respond both rapidly and accurately. Following previous work in our laboratory (Samuel, 1982, 1986), we analyze reaction times in addition to identification shifts. Figure 1 illustrates how these measures reflect adaptation effects.

Insert Figure 1 Here

The top-left panel shows the average identification rate of each of the eight continuum members. The solid curve is from a baseline condition in which the "adaptor" was simply the vowel /a/. This condition produces results that are identical to those obtained with no adaptation, but under timing conditions that are identical to the experimental conditions of interest. The dashed curve shows the identification results after adaptation with the endpoint /ba/. The middle panel on top shows the corresponding comparison between /da/ adaptation and baseline. As is clear, the endpoints produce the usual effect, reducing report of stimuli in the adaptor's category. In the top-right panel, we show the two experimental conditions plotted against each other. As the panel makes clear, adaptation can produce very large changes in identification.

The bottom panels show the corresponding reaction time (RT) results. These are also substantial. The RT effect involves a slowing to responses <u>within the adapted category</u>. We analyze the RT changes by measuring the changes in the overall slope of each RT function. If adaptation with, for example, /ba/ slows down responses within the /ba/ category, this will show up as a "lifting" of the left side of the RT function (the /ba/ end of the test series). For /da/, the "lifting" will be on the right. For the bottom-right panel, this analysis essentially involves adding the amount that the crosses are above the triangles for stimuli 1-3 to the amount that the triangles are above the crosses for stimuli 6-8. This analysis produces an estimate of the RT change, a change of about 58 msec in this example.

Insert Figures 2 & 3 Here

Figures 2 and 3 illustrate the results of combining the laterality manipulation with the reaction time analyses, across the acoustically-based and phonetically-based adaptors. The acoustically-based adaptors were two-formant patterns (F2 and F3). One of these adaptors (F2F3:B) was identical to the endpoint /ba/, except that it had no energy in the first formant frequency region; the other (F2F3:D) was the comparable analog to /da/. These stimuli have little or no phonetic quality (others have called them "nonspeech", but that is probably too strong a claim). These F2F3 adaptors were designed to engage representations that are sensitive to the <u>acoustic</u> structure of /ba/ and /da/. The /pa/ and /ta/ adaptors, as noted, are phonetically quite similar to /ba/ and /da/, respectively, differing only in voicing; the acoustic match is less good. As Figure 2 shows, the F2F3 adaptors were quite effective, producing reliable identification shifts. Moreover, these shifts were accompanied by substantial RT changes, much like those found for the full /ba/ and /da/ endpoints. The figure also makes it clear that these acoustic adaptors are affecting a binaurally-driven level of representation, as both the RT and identification functions are virtually identical for ipsilateral (and binaural) versus contralateral adaptation. Thus, there is a central level of representation that is affected by complex acoustic patterns, which is subject to a slowing of responses through adaptation.

The results for /pa/ and /ta/ adaptation (Figure 3) are quite similar in some ways, and critically different in another. Like the acoustically-matched adaptors, these phonetically-matched adaptors produce reliable identification shifts, and these shifts are clearly produced at a binaurally-driven level. However, as all three bottom panels illustrate, the robust identification shifts are <u>not</u> accompanied by any slowing of responses within the adapted category. This is precisely the sort of qualitative dissociation that is required to posit separate levels of representation for patterns that match acoustically versus those that match phonetically.

The existence of a central level of representation that is sensitive to acoustic matches, and subject to RT changes through adaptation, was rather unexpected. This led us to conduct an additional series of experiments, designed to further pin down the properties of this level (Level II in Table 1). In a set of experiments, we compared adaptation with single formants (either F2 or F3, derived from either /ba/ or /da/) to the effects found for the two-formant F2F3 stimuli. These experiments demonstrated that F2F3 adaptors produce reliably larger effects than the sum of F2 and F3 adaptors: The F2F3 effects <u>cannot</u> be accounted for by a model that relies on the sum of simple effect. Instead, a separate level of representation that is sensitive to an integrative pattern is necessary. The individual F2 (10%, 2 msec) and F3 (7%, 7 msec) adaptors averaged a combined 17% and 9 msec adaptation effect, compared to a 33% F2F3 identification effect, with a 44 msec RT shift. A final experiment in this series tested whether the posited integrative level could combine F2 and F3 information presented in separate ears. In this test, F2 was presented to one ear while F3 went to the other. This dichotic adaptor was compared to a binaural F2F3 presentation mode. In accord with our characterization of Level II as binaurally-driven and operating as an integrative acoustic level, the dichotic adaptor was just as effective as its binaural counterpart, both in terms of the identification shift and the RT change.

The results of our experiments in this series have provided a significant advance in our understanding of the levels of representation that exist in the perceptual processing of speech and other complex sounds. We are continuing this line of investigation, along with several other approaches. The results of these further investigations will be summarized in the Final Technical Report next Spring

IV. LIST OF PUBLICATIONS

Pitt, M. A., and Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. Journal of Experimental Psychology: Human Perception and Performance, 19, 1-27.

Samuel, A. G. (1993). Does lexical information influence the perceptual restoration of phonemes? Submitted to Psychological Review.

Pitt, M. A., and Samuel, A. G. (1993). Lexical and sublexical feedback in auditory word recognition. Submitted to Cognitive Psychology.

Samuel, A. G., and Kat, D. (1993). Levels of Analysis of speech and other complex sounds. Submitted to Cognitive Psychology.

V. PERSONNEL

Principal Investigator: Arthur G. Samuel, Professor of Psychology at the State University of New York at Stony Brook. Ph.D. from University of California, San Diego, 1979.

Senior Research Specialist: Donna Kat, B.A. in Psychology from University of California, San Diego, 1979.

Graduate Student: Lee Wurm. Mr. Wurm joined our lab a year ago, and is conducting research funded by AASERT Grant # 93NL174.

Graduate Student: Sven Mattys. Mr. Mattys joined our lab in September 1993, and is beginning research on lexical issues.

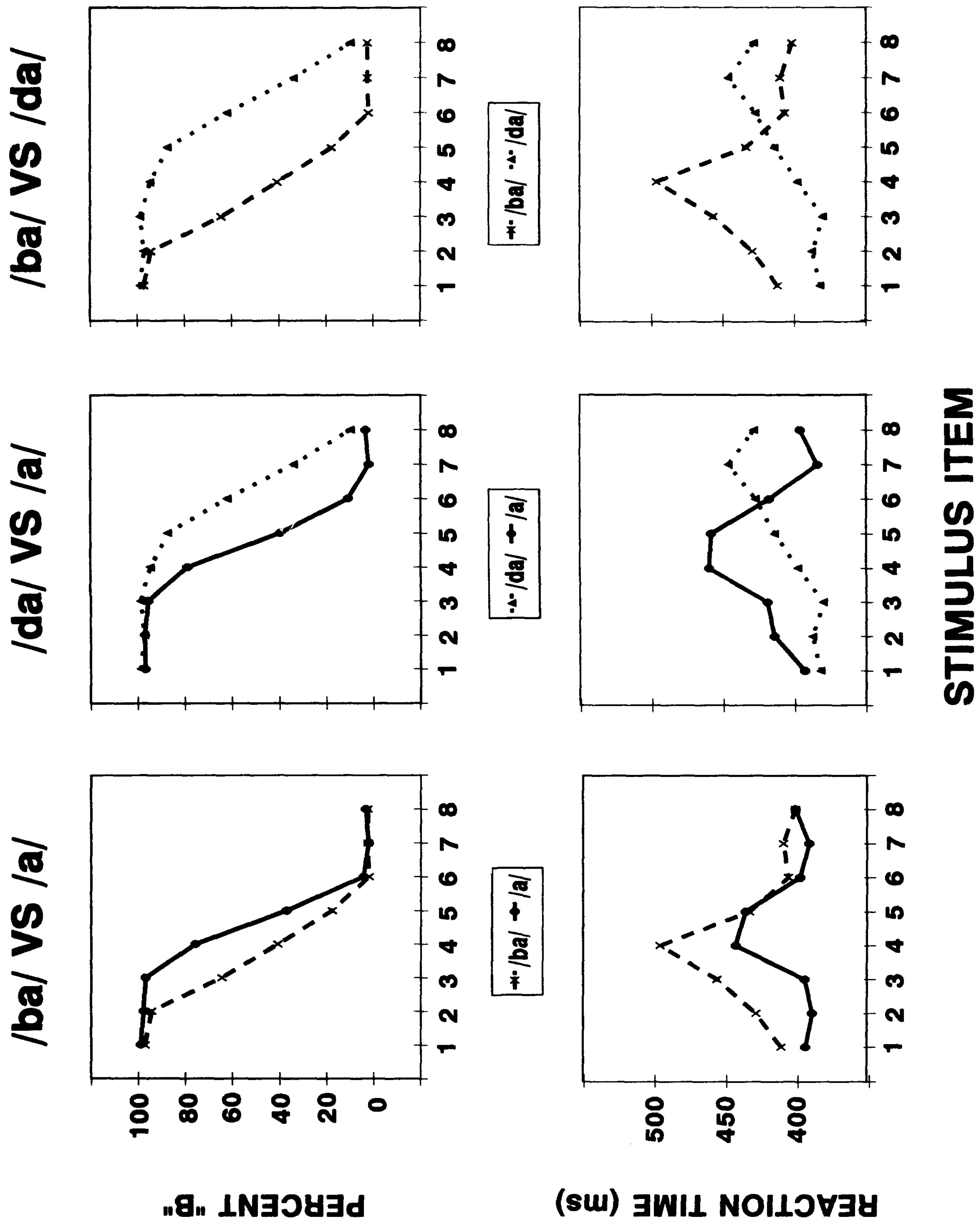
VI. LIST OF INTERACTIONS

The P.I. has been extremely active in national service, serving on the Editorial Boards of leading journals. This past year, these journals were Cognition, Memory and Cognition, and the Journal of Experimental Psychology: Human Perception and Performance. This year, he has agreed to also join the Editorial Board of Perception & Psychophysics. He is also a member of the Perception and Cognition Review Panel for NIMH. These editorial and grant review activities provide a rich source of interaction with top scientists from around the country.

Table 1

Properties	Level I	Level II	Level III
Central versus Peripheral Locus	Peripheral (Monaurally-driven)	Central (Binaurally-driven)	Central (Binaurally-driven)
Nature of Adaptation Shift	Fatigue (?)	Fatigue or Criterion Shift (RT)	Criterion Shift (non-RT)
Stimulus Domain	"Simple" acoustic (Neural Spectrogram)	Integrative Acoustic	Categorical (Phonetic) ?

Figure 1


/ba/ VS /a/
/da/ VS /a/
/ba/ VS /da/
PERCENT "B"
REACTION TIME (ms)
STIMULUS ITEM
/ba/
/a/
/da/
/a/
/ba/
/da/
100
80
60
40
20
0
500
450
400
1 2 3 4 5 6 7 8

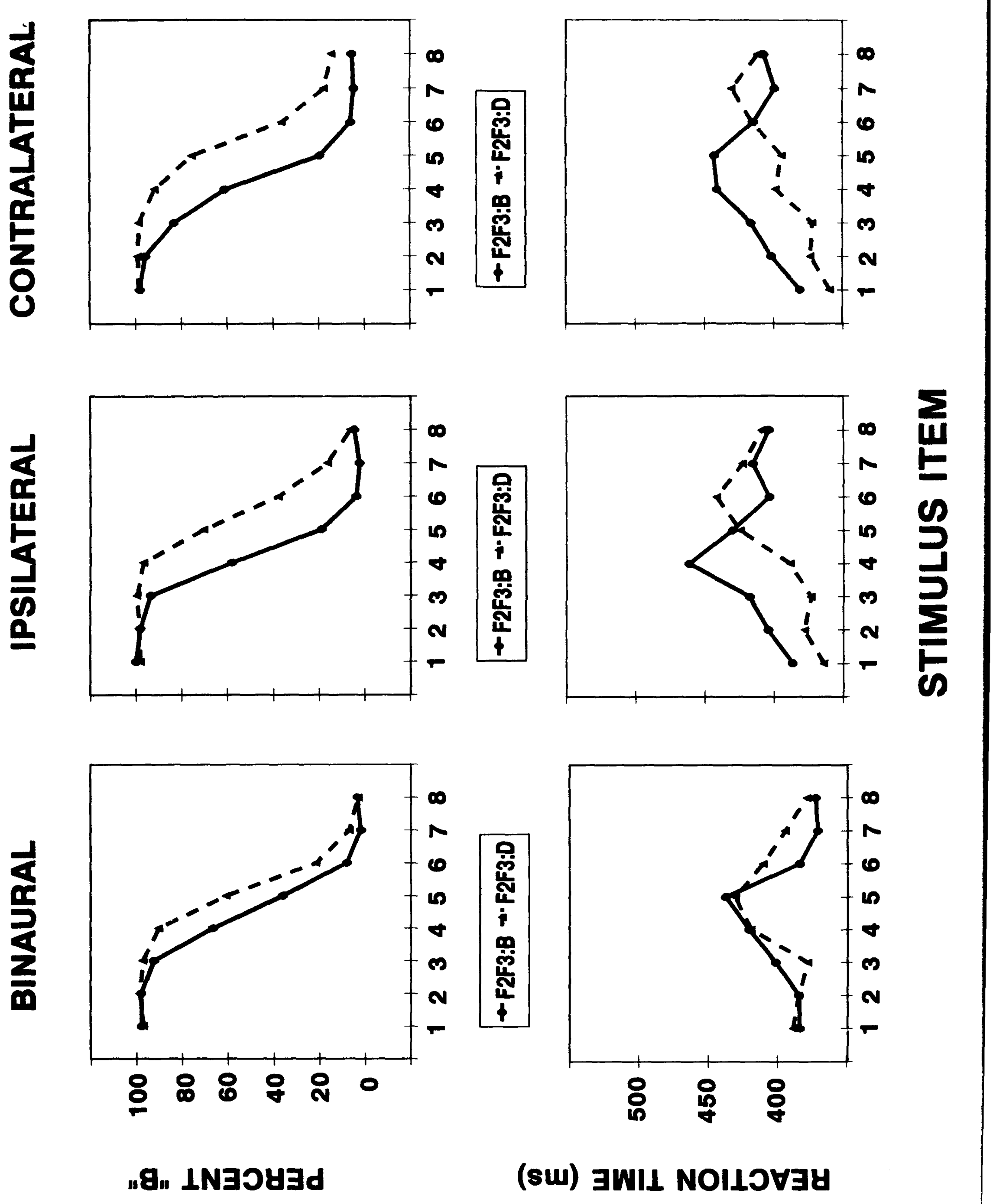
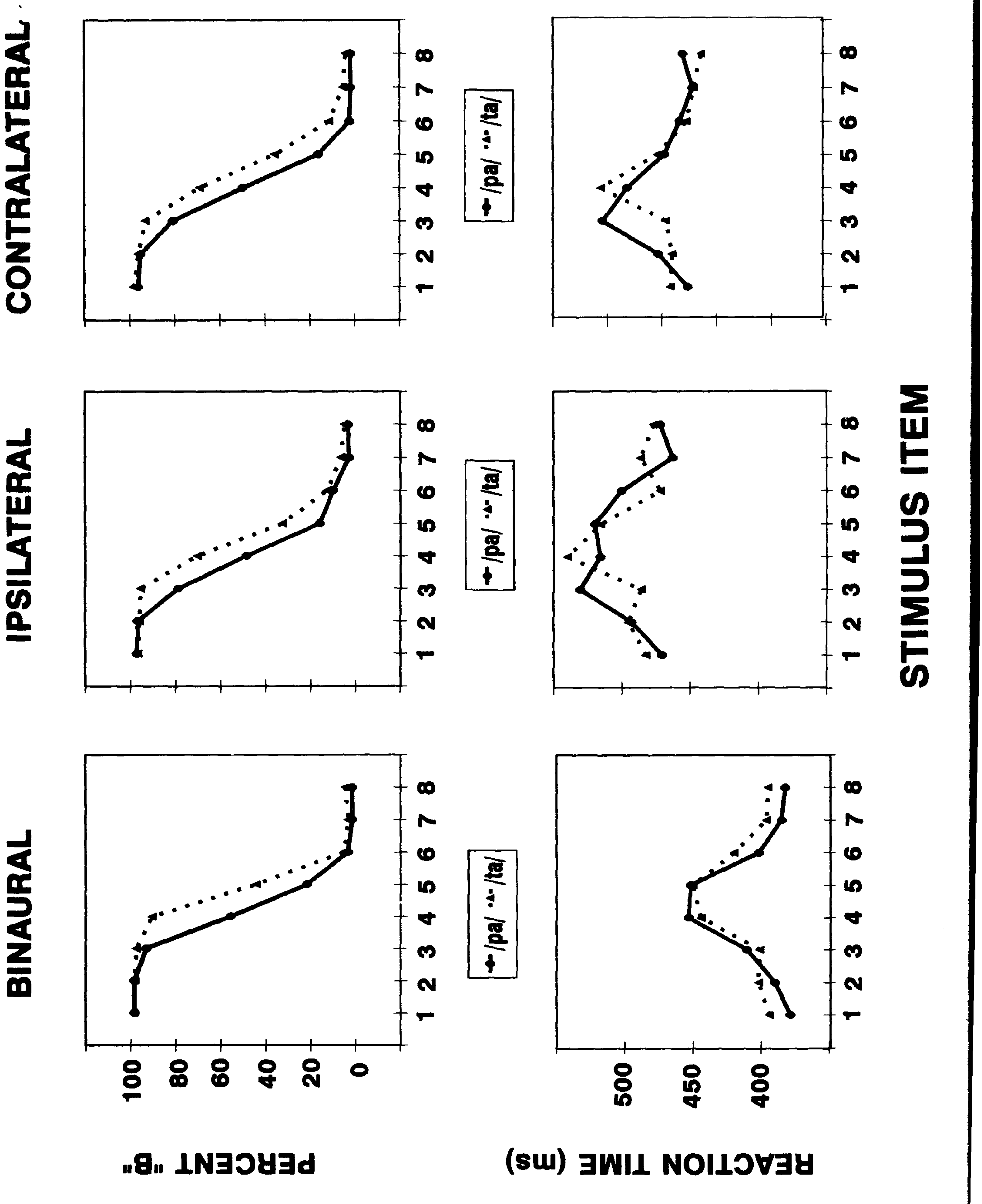

BINAURAL
IPSILATERAL
CONTRALATERAL
PERCENT "B"
100
80
60
40
20
0
REACTION TIME (ms)
500
450
400
F2F3:B
F2F3:D
1 2 3 4 5 6 7 8
STIMULUS ITEM

Figure 3



BINAURAL
IPSILATERAL
CONTRALATERAL
PERCENT "B"
REACTION TIME (ms)
STIMULUS ITEM
/pa/ /ta/